DOCUMENT RESUME

ED 032 776 EM 007 447

By-Anastasio, Ernest J.; And Others

Computer-Assisted Item Writing--II. (Sentence Completion Items).

Educational Testing Service, Princeton, N.J.
Report No-TDM-69-1
Pub Date Jun 69
Note-13p.; Paper presented at National Council on Measurement in Education (Annual, Los Angeles, California, Feb. 1969)
EDRS Price MF-$0.25 HC-$0.75
Descriptors-Computational Linguistics, Computer Oriented Programs, Computer Programs, Educational Research, *Item Analysis, Linguistics, Machine Translation, Multiple Choice Tests, Programing, Semantics, Sentence Diagraming, Sentence Structure, Structural Analysis, *Test Construction, Testing, Test Selection

A study to facilitate computer generation of test items succeeded in developing a set of criteria which could be used by a computer to select sentences suitable for sentence completion question items. The sentence completion item requires the student to complete a sentence from which one or more words have been deleted. Analysis of 110 appropriate test sentences showed that they often could be characterized by sentence length, internal punctuation, and use of conjunctions and prepositional phrases. These criteria were used (without writing a computer program) to analyze a magazine essay, and the sentences thus selected were submitted to experienced sentence completion writers who found them suitable for use. Further research will involve the actual writing of the program to select appropriate sentences as well as an attempt to develop a computer ability to generate its own test items. (BB)

ED0 32776

TEST DEVELOPMENT MEMORANDUM
TDM-69-1 JUNE 1969

Computer-Assisted Item Writing—II (Sentence Completion Items)

Ernest J. Anastasio
Developmental Research Division, ETS
Donald M. Marcotte
Wayne State University
John Fremer
Test Development Division, ETS

EM007447

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY
BERKELEY, CALIFORNIA

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

COMPUTER-ASSISTED ITEM WRITING — II (SENTENCE COMPLETION ITEMS)

ERNEST J. ANASTASIO
Developmental Research Division, ETS

DONALD M. MARCOTTE
Wayne State University

JOHN FREMER
Test Development Division, ETS

Test Development Memorandum
TDM-69-1
June 1969

EDUCATIONAL TESTING SERVICE
Princeton, New Jersey

Computer-Assisted Item Writing — II (Sentence Completion Items)[1]

General Introduction

In recent months, a number of researchers have undertaken projects designed to exploit the potential of computers for testing. For example, computer technicians and test specialists have made great progress in developing and implementing systems to turn over to the computer a major portion of the task of assembling appropriately classified and pretested items into tests. Also, the potential for computer assistance in the preparation of test items has recently been demonstrated by researchers (e.g., Richards, 1967; Fremer & Anastasio, 1969) who, after systematic analyses of discrete verbal items, have shown that it is possible for the computer to carry out some of the less creative portions of item writing.

This report is the second in a continuing study of computer assistance in the writing of test items. The main objective of the project is to identify the properties which characterize "good" test items for measuring verbal ability and to develop rules for coding words and sentences so that they can be manipulated by a computer. Ultimately, such systematic exploration of the item-writing process may make possible computer generation of test items with prespecified properties,

---

[1]Paper presented at the annual meeting of the National Council on Measurement in Education, Los Angeles, California, February 1969.

that is, items that would require only minor editing before preliminary field testing. Whether or not this outcome is achieved, the research should increase our understanding of the structure and functioning of verbal ability items.

The initial study in this series focused on spelling items, which were studied first because they seemed to have the fewest facets or dimensions. In that study an analysis was made of the types of misspellings used by writers of spelling items and a set of error-generation rules, in the form of a computer program, was developed. The program was then used to analyze a sample of English words; and a list of misspelled words, separated into previously defined error categories, was created. Subsequently, the list was evaluated by spelling-test developers and judged to be a useful resource.

Sentence Completion Items

In the current study a more complex item type, the sentence completion item, was considered. This type of item requires the student to complete a sentence from which one or more words have been deleted. He must identify, from a list of five choices, the word or set of words which would make the best sense when inserted in the sentence. Such questions provide a measure of one aspect of English comprehension: the ability to recognize logical and stylistic consistency among the elements in a sentence. If a student understands the implications of a sentence, he should be able to select the answer that not only best fulfills the

meaning of the sentence, but also completes it in the most lucid, logical, and stylistically consistent manner. Figure 1 shows directions for answering sentence completion items, and includes a sample item appropriate for college freshmen.

Figure 1

Directions: Each of the sentences below has one or more blank spaces, each blank indicating that a word has been omitted. Beneath the sentence are five lettered words or sets of words. You are to choose the one word or set of words which, when inserted in the sentence, best fits in with the meaning of the sentence as a whole.

Sample Item

Aggression need not be -------; it may be a learned way of ------- a problem.

(A) furtive..evading
(B) hostile..solving
(C) immediate..escaping
(D) antisocial..comprehending
(E) prolonged..postponing

The task of writing sentence completion items can be briefly outlined as follows:

1. Selecting an appropriate sentence (this involves thumbing through newspapers, magazines, and books)

2. Choosing the words to be deleted

3. Revising the sentence, if necessary, to make the task of writing distracters easier

4. Choosing appropriate distracters or incorrect options

After an item writer at Educational Testing Service (ETS) has polished a sentence completion item to his satisfaction, the item

is submitted to a series of at least seven reviews, which may or may not result in the item's being tried in a pretest. A typical reason for rejecting a prospective sentence completion item is that the sentence lacks sufficient redundancy to determine a single key.

The following principles are currently being used by item writers at ETS in the selection of sentences for sentence completion items for use with college level students.

Sentences should be (1) grammatically correct; (2) interdependent; and (3) often compound and complex. Sentences should suggest an idea of balance, comparison, contrast, or antithesis; or sentences should present a restatement or illustration of a main idea.

These selection principles contain only one suggestion regarding the syntactic structure of an appropriate sentence—that it be compound or complex. After examining a collection of "good" sentence completion items (items that had passed through the review process, had been pretested, and had proved to have high discrimination indices), the authors discovered considerable structural similarity. It seemed clear that, if it were possible to identify particular structures or other descriptive or enumerative features that were consistently found in "good" sentence completion items, the amount of effort expended in the selection of sentences could be reduced considerably. If a sufficiently objective set of rules were developed, a clerical assistant might be able to conduct a search. If appropriately classified textual material were available, a computer could carry out the search. In this project we

would satisfy our goal even if the computer could be used only to identify some subset of the total range of sentences which were presently being developed.

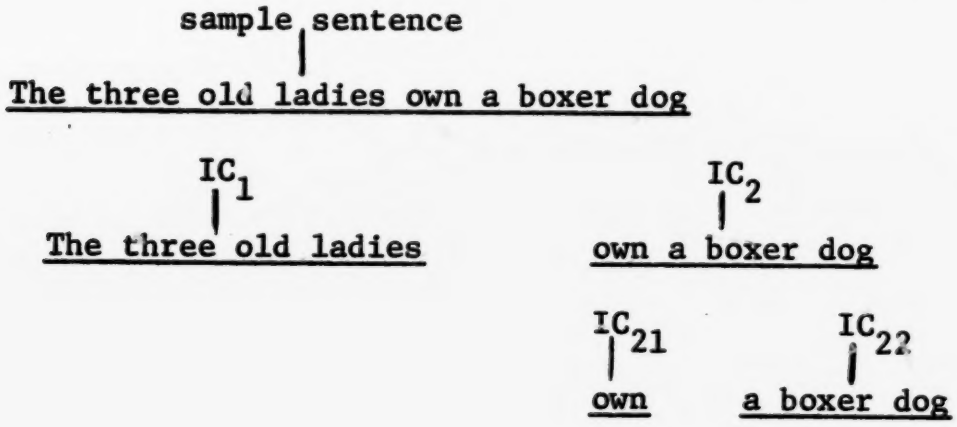
For the present study then we were concerned solely with exploring the characteristics of sentences currently used, with a view towards developing sentence selection rules. This study did not attempt to handle the issue of sentence generation—a problem which involves lexical as well as syntactical considerations.

Method of Sentence Analysis

An analysis was carried out on a total of 110 "good" sentence completion items. First, a formal syntactic analysis was applied to the items to obtain descriptions of the syntactic relations in each sentence. Francis' "Chinese box" scheme—an immediate constituents technique—was used because of the simplicity of its categorization system. Immediate constituents analysis operates on the assumption that most sentences can be divided initially into two parts, each of which may be further divided. Essentially, this analysis isolates the subject, verb, and object of a sentence. A simplified example of the technique appears in Figure 2.

Figure 2

Example of Immediate Constituents (IC) Analysis
(adapted from Gleason, Linguistics and English Grammar, 1965)



In the sample sentence in Figure 2, The three old ladies own a boxer dog (adapted from Gleason, 1965), The three old ladies and own a boxer dog are called the immediate constituents of the sentence. In turn, each of these can be divided into its immediate constituents. In the example, only the second constituent is divided, into own and a boxer dog. Note that three major sentence elements have been obtained, but they have not been obtained as coordinate parts of the sentence. At each stage of the analysis, only the immediate constituents—that is, the immediate components of the construction—have been considered. In the next step of the analysis, constructions within each constituent can be analyzed—for example, pairs of words like old ladies and boxer dog.

For the immediate constituents analysis, frequency data were collected on each of 29 variables listed in Figure 3.

Figure 3

Variables Used in the Analysis of the Sentences

1. Blank arrangement
2. Number of words in the sentence
3. Pattern of phrases and clauses (pattern complexity)
4. Spatial relation of blank or blanks
5. Spatial relation of blank or blanks according to pattern
6. Number of prepositional phrases
7. Number of terminal prepositional phrases
8. Part of speech of blank within a terminal prepositional phrase
9. Part of speech of blank within a nonterminal prepositional phrase
10. Distribution of prepositional phrases according to pattern
11. Syntactic structure of blank
12. Part of speech of blank
13. Subject-verb separation
14. Inverted subject verb
15. Verb separation (if auxilaries used)
16. Punctuation
17. Negation
18. Coordination
19. Balance (number of words) of coordinated elements
20. Subject or subjects (sirgular or plural)
21. Infinitives
22. Infinitives with auxilaries
23. First-order includers
24. Blank within first-order includer
25. Second-order includers
26. Blank within second-order includer
27. Substitute noun for subject
28. Spatial separation between subject and substitute noun
29. Use of quotation marks

Inspection of the frequency data indicated that 80 per cent of the items could be characterized by the following variables:

1. Sentence length between 18 and 28 words
2. Presence of internal punctuation—comma, semicolon, or colon
3. Use of a subordinating and/or coordinating conjunction
4. At least 2 prepositional phrases

5. At least 1 terminal prepositional phrase in a dependent or independent clause

With the exception of variable 5, each of these variables could easily be incorporated into a computer program for selecting sentences. Word counting (variable 1) is easily accomplished—words being defined as any group of nonspecial characters which fall between two blanks. Identification of internal punctuation, subordinating or coordinating conjunctions, and prepositions (variables 2-4) is easily accomplished by comparing words and special characters against master lists. Handling variable 5 would be more difficult, however. If terminal prepositional phrases are operationally defined as any group of two-five words headed by a preposition and terminated with internal or terminal punctuation, then, they too can be recognized by a computer. In carrying out this operation, however, the computer would default in recognizing terminal prepositional phrases in restrictive clauses, which are not marked off by any punctuation.

Because a suitable data base was not available, a computer program was not actually written. Rather, a computer program was simulated, and the operations were carried out by an individual who was totally naïve with respect to this research. The individual was presented with a list of the rules, a master lister "dictionary" of prepositions and conjunctions, and a randomly selected issue of the Saturday Review. The instructions were to select an essay from the magazine and systematically apply the rules to each sentence in the essay. If a match was found on 4 of the 5 variables on the list, the instructions were to write out

that sentence on a separate piece of paper. A total of 26 sentences were identified out of a sample of approximately 300 sentences.

Evaluation

The suitability of the sentences selected by the simulated program was clearly a critical question. Would these selected sentences be judged by test developers as potentially usable material for sentence completion items? It was decided that the selected sentences would be judged suitable if they could successfully pass review by a group of experienced item writers.

Sixteen sentences were randomly eliminated from the list of 26 clerically selected items, and 16 pretested items that had previously been written, and reviewed, by human item writers, were put in their place. The 26 sentences were then submitted for review to each of seven item writers. Instructions were to make a yes or no judgment about the potential, as a sentence completion item, of each of the sentences on the list.

No meaningful differences were found between the number of sentences chosen from each group (chi-square with one degree of freedom). Clerically selected sentences were judged to be appropriate for sentence completion items equally as often as sentences that had already been used in sentence completion items. Thus we would seem to have satisfied our central objective.

As a consideration for future research, further elaboration of the sentence selection rules and investigation of the relative usefulness of various sources of sentences is clearly required. Also, the clerically selected sentences from various sources should be traced through a complete cycle of item writing, reviewing, and pretesting. Plans are now being formulated to pursue these two lines of research.

References

Fremer, J. & Anastasio, E. J. Computer-assisted item writing — I (spelling items). Journal of Educational Measurement, 1969, 6, in press.

Gleason, H. A., Jr. Linguistics and English grammar, New York: Holt, Rinehart and Winston, 1965.

Richards, J. M., Jr. Can computers write college admissions tests? Journal of Applied Psychology, 1967, 51, 211-215.